ASCI Terascale Scientific Data Analysis and Visualization

Objective

The ASCI visualization effort is developing advanced data structures and algorithms for the analysis and visualization of terascale scientific data in a distributed-data environment. Hierarchical data representations and algorithms for progressive computation and rendering play a key role in the design of a layered architecture that distributes data and computation across the range of available resources.

Impact

Comprehensive data analysis tools are critical in assessing the usefulness of the massive volumes of data that will be generated by ASCI and other large-scale scientific simulation codes.

The Accelerated Strategic Computing Initiative (ASCI) vision is to shift rapidly from test-based nuclear stockpile stewardship to a high-performance computing and data analysis environment that demonstrates unprecedented performance, reliability, and confidence in numerical simulation. ASCI applications are using extremely high fidelity computer models (on the order of one billion cells) to generate terabytes of raw data. These data are being analyzed by physicists who rely on the visualization capabilities being developed within the ASCI program. The challenge of analyzing and visualizing the terascale datasets is addressed by combining high-performance storage and networking with a scalable visualization architecture that permits interactive exploration of large volumes of data.

Visualization Architecture

Terascale visualization architecture is driven by the need to visualize the terabytes of data resulting from simulations that may have been computed thousands of miles away. Flexibility and adaptability of data access mechanisms are required to accommodate orders-of-magnitude variations in processing, memory, and storage. Our overall strategy for terascale visualization has three components: Prepare, Browse, and Select. Preparation refers to the processing that is performed on the resulting simulation data to provide the required flexibility and adaptability in data access. Browsing is a term used for high-performance interaction with the prepared data. Typically, few visualization modes are supported at high performance, and hence browsing is intended primarily to be a method of initial data exploration to find features of interest. Selection occurs when critical features are discovered and the user requires a more quantitative understanding of what is happening in a particular space-time neighborhood. The data of interest are then communicated to a full-service visualization tool. This approach allows us to leverage the tremendous amount of effort expended in building visualization and analysis tools that typically do not scale to the terabyte range of data being considered.

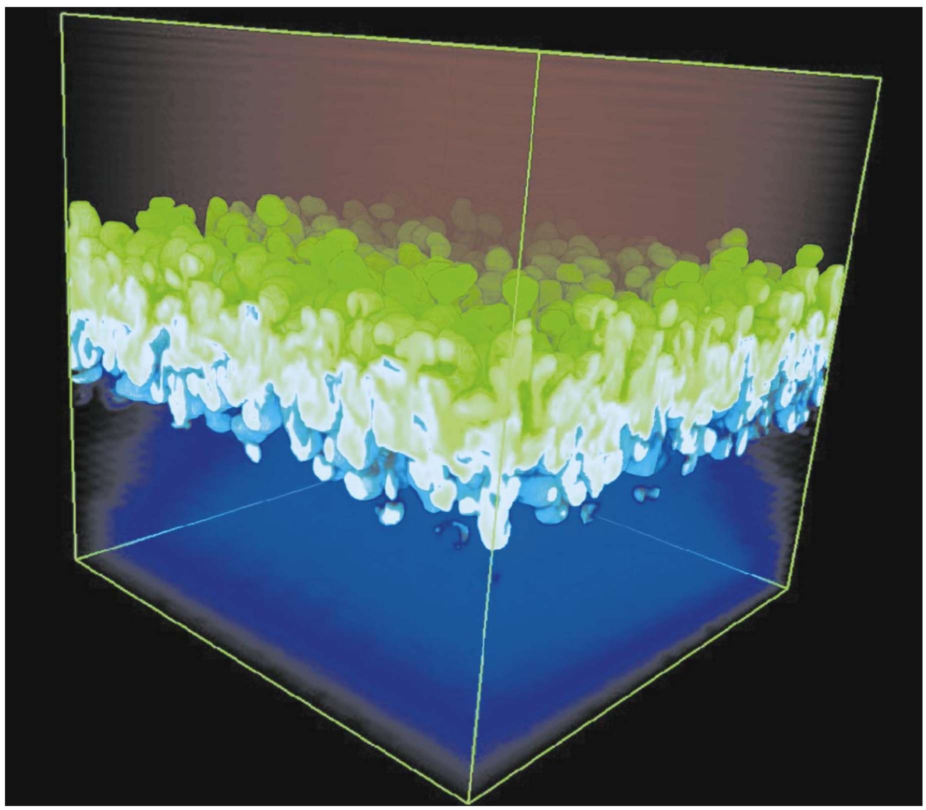

Figure 1. Volume rendering of a Raleigh-Taylor instability simulation in the browser. The compressed volume data is paged from disk and decompressed transparently.

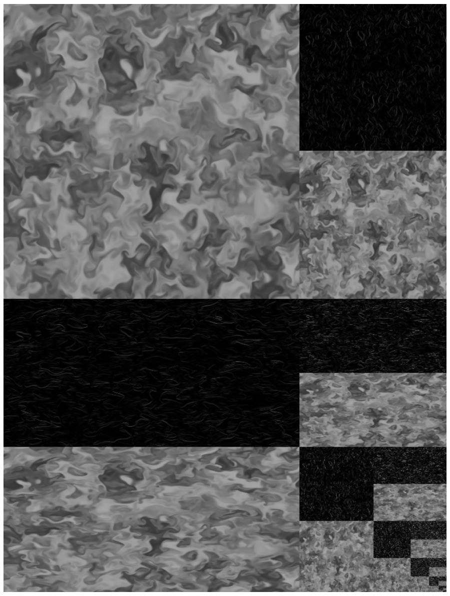

Figure 2. Illustration of the application of a novel wavelet transform to a 2-D slice of data from a Raleigh-Taylor simulation.

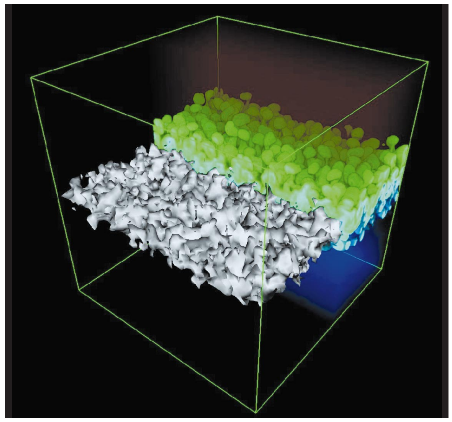

Figure 3. Isosurface and volume rendering parameters are interactively manipulated simulataneously in the browser.

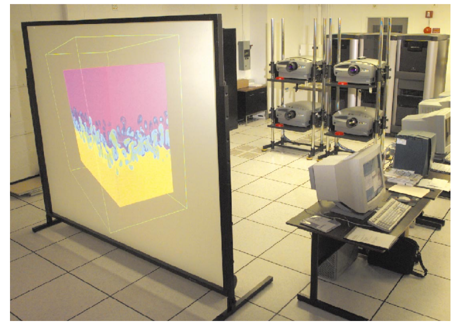

Figure 4. High-resolution display technologies are being investigated. The four-projector display wall shown here has resolution 2560 x 2048, and can be used both for interactive applications and precomputed animations.

Data Services

Intelligent data storage and caching are key to the ability to interact with terascale data. Conceptually, the data server lies between the simulation code and the visualization application, providing access to the portions of the data required for processing. The principal responsibilities of the data server are to satisfy client queries for sub-meshes and slices, both at full and adaptive resolutions. Hierarchical data representations permit spatial queries and selective refinement based on preprocessed error metrics, which are stored in the hierarchy. The data server may be located on the MPP running a simulation, or on an auxiliary server or a cluster of workstations with large aggregate memory and access to high-performance storage.

In order to use available memory and storage efficiently, we are evaluating approaches to caching compressed data in primary memory, in addition to paging and high-speed transmission of compressed representations for distance visualization. Research is currently being conducted in 3-D and 4-D wavelet compression with adaptive sparse decompression and paging capabilities. Figure 2 illustrates a novel wavelet transform that has been developed for lossless compression of quantized scientific data.

Visual Browser

The data browser allows interactive navigation of terascale scientific data by acting as a client of the data server. The user interactively manipulates parameters for isovalues, orthogonal slices, and volume images while gaining a global view of the simulation's time dynamics through an intuitive interface for navigating the time dimension. The browser may be integrated into existing visualization systems in order to leverage the comprehensive visualization tools available. Figure 3 illustrates the volume and surface visualization capabilities of the browser.

Interactive Techniques

Scalable visualization requires efficient algorithmic techniques for accessing only the data required for a particular visualization.

Preprocessing of hierarchical data representations allows the computation of view-dependent error estimates from the hierarchical error bounds on the data. We are using these view-dependent error measures at all stages of the visualization pipeline, reducing not only the rendering requirements, but also improving utilization of processors, memory, storage, and networks.

Display Technologies

Terascale data visualization introduces new demands on the availability of display resolution. Calculations are both highly adaptive and have a very high spatial density of samples, which leads to the need for display resolutions beyond the capability of traditional desktop displays. Figure 4 shows a prototype display wall capability. The current display capability is 2560 x 2048 resolution using four projectors, with plans for a production capability of resolution 3840 x 3072 with nine projectors by the end of 1998. We are currently investigating the software architecture changes that are required for efficiently rendering on such large displays. Such displays are typically constructed by aggregating multiple high-end graphics engines in one or more high-performance server or workstation.

For more information on ASCI Visualization efforts at LLNL, contact:
Mark Duchaineau, (925) 423-1320, duchaine@llnl.gov or
Dan Schikore, (925) 424-5799, schikore@llnl.gov.

ASCI, the Accelerated Strategic Computing Initiative Program, includes participants from Los Alamos National Laboratory, Lawrence Livermore National Laboratory, and Sandia National Laboratories. Work performed under the auspices of the U.S. DOE by LLNL under contract No. W-7405-ENG-48.